Machine translation, translation errors, and adequacy: Spanish-English vs. Spanish-Romanian

Laura Monguilod
lauramong@hotmail.com
Bianca Vitalaru
bianca.vitalaru@uah.es
Universidad de Alcalá

Abstract

This paper has two objectives: 1. To analyse the adequacy of using neural machine translation (NMT) for the translation of health information (from Spanish into English and Romanian) used in Spanish public health campaigns; and 2. To compare results considering these two linguistic combinations. Results show that post-editing is essential to improve the quality of the translations for both language combinations since they cannot be used as a primary resource for informing foreign users without post-editing. Moreover, Romanian translations require more post-editing. However, using NMT for informative texts combined with human post-editing can be used as a strategy to benefit from the potential of MT while at the same time ensuring the quality of the public service translations depending on the language combination and on the amount of time allotted for the task.

1 Introduction

Within the context of globalisation and crisis situations characterised by the increase of the migrants' percentage in Europe, Spain's foreign population kept growing in the last decade. In fact, it increased by 182,141 people in the first half of 2022 (INE, 2022). This situation has affected both the general needs for social integration and access to public services (i.e., education, administration, healthcare, or social welfare) and specifically the communication needs to access these rights in several language pairs. On the other hand, as Navaza, Estévez, and Serrano (2009) underline, healthcare providers who needed to be able to inform patients who did not understand Spanish were also affected. One of the foreign populations that maintained its percentage in the last decade in Spain is the Romanian population (See, for example, that it had more than 600.000 people every year since 2008 according to INE, 1998-2022).

In the specific context of the healthcare settings, as one of the rights that the foreign population has, health campaigns are used as a tool to transmit essential information regarding healthcare to the general population, usually focusing on disease prevention or situations that can imply some level of risk. They also include or even specifically address the foreign population, who, due to communication difficulties or cultural differences, may not be aware of the risks they are exposed to and of the prevention measures that they should take. In fact, both linguistic and cultural differences can hinder the dissemination of healthcare-related materials (Sixsmith et al., 2014).

On the other hand, technology in general can assist in the provision of (public) services to both the general and the foreign population and facilitates the provision of services that were previously unavailable (Sánchez Ramos & Rico Pérez, 2020). Technology is also fundamental in the translation and interpreting (T&I) sector,

Proceedings of the First NLP4TIA Workshop, pages 4–12
Varna, Sep 8, 2023
https://doi.org/10.26615/978-954-452-091-5_002

which is the channel that facilitates communication when the foreign population is involved. This channel is specifically relevant within the Public Service Interpreting and Translation (PSIT) context, that is, fields such as education, administration, social welfare, healthcare, and legal settings (Sánchez Ramos & Rico Pérez, 2020; Valero-Garcés, 2018), to which the foreign population has free access. Additionally, within the private T&I sector, technologies (especially CAT tools and including Machine Translation (MT) and post-editing) are an essential part of translators' daily work to reduce costs, increase efficiency, and improve productivity (Sánchez Ramos & Rico Pérez, 2020). Moreover, if an effective post-editing process is followed, the use of MT or terminology management tools can improve productivity, assist with textual consistency, and ensure better quality (Sánchez Ramos & Rico Pérez, 2020). Furthermore, some other advantages should be considered, such as the fact that CAT tools allow the translator to store their work so that they can use it again when they need to work with similar texts (Kerremans et al., 2018). However, the use of technology (especially MT), also generates debates among professionals regarding the quality of the product obtained due to "terminological inconsistencies, false meanings, and a clear lack of syntactic and stylistic systematicity" (Kerremans et al., 2018). Finally, specialists have also been reluctant to incorporate translation tools in migrant support contexts (Sánchez Ramos & Rico Pérez, 2020).

Considering this context, this paper has two objectives:

1. To analyse the adequacy of using neural machine translation (NMT) for the translation of health information (from Spanish into English and Romanian) used in Spanish public health campaigns.
2. To compare results considering the Spanish-English (ES-EN) and the Spanish-Romanian (ES-RO) combinations.

We based this study on two hypotheses: 1) NMT outputs are not completely adequate if used as a primary resource for informing foreign users without post-editing and 2) there will be more translation errors in the Spanish-Romanian combination than in the Spanish-English combination, which will, in turn, require more post-editing.

2 Classifying translation errors: human and machine translation

A translation error can be defined as an inappropriate equivalence (Hurtado Albir, 2011). Translation errors are directly related to translation problems (Hurtado Albir, 2011) especially because translation problems are seldom found in the translation process (Gregorio Cano, 2017) and can be identified in advance. In addition, they are different from translation difficulties since they do not depend on the translator's ability to solve problems as an individual (Nord, 2007).

Translation errors are fundamental when analysing and evaluating translation quality and can be classified based on different criteria. The most frequent categories one can find when defining and analysing translation errors are related to errors concerning either the source text or the target language and the two main phases of translation: comprehension and re-expression (Hurtado Albir, 2011). Moreover, the type of translator (human or MT) is another criterion to be considered. In fact, we believe that to identify the types of errors and analyse them correctly we need to be aware of the similarities and differences between the outputs of both types of translators. That is why we show, in Table 1, a list and a basic comparison between two classifications of translation errors:

Common human translation errors (Delisle, 1993, cited in Hurtado Albir, 2011)	**Common MT errors** (Alarcón Navío, 2003)
False sense.	False meanings.
Countermeaning.	Nonsenses.
Nonsense.	Terminological improprieties.
	Syntactic and lexical calques of the source language.
Addition.	-
Omission.	Untranslated words.
Hypertranslation.	Repetitions.
Overtranslation.	Unnecessary foreign words.
Undertranslation.	Alteration of word order and punctuation.
	Incorrect use of prepositions and verb tenses.
	Incorrect translation of double negation.
	Errors in the translation of lexicalised metaphors.

Table 1: Classification of translation errors

Considering that this research focuses on the use of NMT, we will provide the characteristics for the types of MT errors whose meaning might require some clarification (as stated by Alarcón Navío, 2003) and the list that will be used to identify translation errors in our study:

- False meanings: hinder text comprehension by choosing a term that can be considered similar but is incorrect.

- Nonsenses: a consequence of discursive incoherence and syntactic structures that are difficult to understand. According to Vázquez and del Arbol (2008), two types of nonsense are particularly significant: on the one hand, [simple] nonsense, understood as a mistake that can hinder text comprehension, and complete nonsense words/expressions, understood as a type of nonsense that renders the discourse meaningless and illogical.

- Terminological improprieties: the target term that has been chosen is not quite adequate since the most general definition of the term has been used.

- Syntactic and lexical calques of the source language: borrowing word order and structures from the source language.

- Untranslated words: using the source terms instead of translations.

- Unnecessary foreign words: using unnecessary loans.

- Others: repetitions, alterations of word order or punctuation, incorrect use of prepositions or verb tenses, incorrect translation of double negations, errors in the translation of lexicalised metaphors.

3 Methodology

The methodology used to obtain information is descriptive and it involves corpus compilation, error tagging using MT and the Raw Output Evaluator tool, as well as counting and analysing translation errors.

Specifically, we took the following steps:

1. Corpus compilation of texts from three health campaigns of the Spanish Ministry of Health (see Table 2). The texts were chosen considering the importance of healthcare campaigns for prevention and healthcare purposes in general and their specific role when it comes to informing the foreign population on disease prevention or situations that imply some level of risk. Ultimately, we also kept in mind the fact that linguistic and cultural differences have been found to hinder the distribution of healthcare-related materials (Álvaro Aranda, 2020).

Campaign	**Nº of words**
1. Malos Humos	206
2. Alcohol	148
3. Conducta suicida	185

Table 2: Corpus

2. Feeding the texts of the campaigns to two free online MT engines: DeepL and Google Translator (GT) and using the translations to create a translations corpus for each language (English and Romanian). Free online MT engines were chosen for two main reasons: the fact that free tools are available and widely used nowadays, especially by trainees and the current scarcity or even lack of funds to provide PSIT services in Spain. From this point of view, our basic analysis of the situation showed that associations, NGO, and organisations that usually provide social or economic assistance to the foreign population usually rely on little financial help regarding linguistic assistance. Thus, they need to turn to free tools. By analysing these engines, we can have an insight into the MT that can be used by organisations for which these MT are the only available options and the contexts in which they could be used.

3. Uploading the translations corpora to the Raw Output Evaluator tool.

4. Identifying and tagging translation errors in the corpora of translations available within the Raw Output Evaluator tool. To determine the errors, we applied Alarcón Navío's (2003) classification of common MT errors (see Table 1). The Raw Output Evaluator is a tool that helps the user to compare several translations at once. These translations can either be generated within the tool or can be uploaded to the tool and the types of errors must be manually tagged by the user. The tool can also be

used during the post-editing process, and it allows the classification of the different types of errors found when using MT (Farrell, 2018).

5. Analysing the types of errors in context and counting the number of errors.

6. Evaluating the adequacy of the outputs considering the types of translation errors previously established and the percentage of errors. We relied on Reiss's (1983) definition of adequacy, which is based on "appropriateness". We first considered that the translation was adequate when it was considered appropriate in a specific medical, social, and cultural context/setting. Additionally, we also kept in mind the number of translation errors found in relation to the total number of words of the corpus compiled.

7. Comparing the results considering the English and the Romanian corpora.

4 Results

We chose three campaigns to test our hypotheses and show our results and several examples of translation errors have been included by type of error in Table 3 and Table 4.

MT errors	**Original text (ES)**	**DeepL (EN)**	**GT (EN)**
False meanings	*Disfrutar de un entorno saludable es primordial. No fumar tabaco ni relacionados lo hace posible, por eso di ¿MALOS HUMOS? NO, GRACIAS*	Enjoying a healthy environment is paramount. Not smoking tobacco or related makes it possible, that's why I said BAD SMOKES? NO, THANKS	Enjoying a healthy environment is paramount. Not smoking tobacco or related products makes it possible, so say NO SMOKE? NO, THANK YOU
Syntactic and lexical calques of the source languages	*Disfrutar de un entorno saludable es primordial. No fumar tabaco ni relacionados lo hace posible, por eso di ¿MALOS HUMOS? NO, GRACIAS*	Enjoying a healthy environment is paramount. Not smoking tobacco or related makes it possible, that's why I said BAD SMOKES? NO, THANKS	Enjoying a healthy environment is paramount. Not smoking tobacco or related products makes it possible, so say NO SMOKE? NO, THANK YOU
Terminological improprieties/imprecise expressions	*La mejor opción es dejarlo, y si no lo has hecho aún, respeta en esos espacios a los demás, especialmente a personas vulnerables.*	The best option is to leave it, and if you haven't done it yet, respect others in those spaces, especially vulnerable people.	The best option is to quit, and if you have not already done so, respect others in these spaces, especially vulnerable people.
Repetitions	*Estudios recientes señalan que un número creciente de hombres gais, bisexuales y otros hombres que tienen relaciones sexuales con hombres (GBHSH) tiene el VIH.*	Recent studies indicate that a growing number of gay, bisexual, and other men who have sex with men (GBHSM) men have HIV.	Recent studies indicate that a growing number of gay, bisexual and other men who have sex with men (GBHSH) have HIV.
Lexicalised metaphors	*EL HUMO NO TE DEJA VER.*	THE SMOKE DOES NOT LET YOU SEE.	SMOKE IS BLIND.

Table 3: Examples of the types of errors found in the English translations

MT errors	**Original text (ES) [authors' English translation]**	**DeepL (RO) [authors' English translation]**	**GT (RO)**

Syntactic and lexical calques of the source languages False meanings	*El Ministerio de Sanidad promueve la Línea 024 de atención a la conducta suicida* [The Ministry of Health promotes the 024-Suicide line to assist people with suicidal behaviours].	Ministerul Sănătății promovează linia telefonică 024 pentru suicid [The Ministry of Health promotes the 024-telephone line for suicide].	Ministerul Sănătății promovează Linia fierbinte 024 pentru atenția asupra comportamentului suicidar. [The Ministry of Health promotes the 024 Hotline for attention towards suicidal behaviour].
Terminological improprieties/imprecise expressions	*Somos uno de los países con más vacunados: el 93% de la población mayor de 12 años se ha administrado la pauta completa de primovacunación* [We are one of the countries with the highest number of vaccinated people: 93% of the population over the age of 12 has received the full primary vaccination series].	Suntem una dintre cele mai vaccinate țări din lume: 93% din populația cu vârsta de peste 12 ani a primit vaccinarea primară completă [We are one of the world's most vaccinated countries: 93% of the population over the age of 12 has received the full primary vaccination].	Suntem una dintre țările cu cele mai vaccinate: 93% din populația cu vârsta peste 12 ani a primit programul complet de vaccinare primară [We are one of the countries with the most vaccinated*: 93% of the population over the age of 12 has received the full primary vaccination programme].
	Lo importante es no participar [The important thing is to not participate].	Este important să nu participați [It is important not to participate].	Important este să nu participi [What matters is that you do not participate].

Table 4: Examples of errors found in the Romanian translations

Four aspects were particularly significant. First, we found significant examples in terms of the changing of the meaning as a result of syntactic and lexical calques and false meanings: e.g., the literal translation for "El Ministerio de Sanidad promueve la Línea 024 de atención a la conducta suicida" was 1) "Ministerul Sănătății promovează linia telefonică 024 pentru suicid" [translation of the Romanian version: "The Ministry of Health promotes the telephone line 024 for suicide"] (DeepL) and 2) "Ministerul Sănătății promovează Linia fierbinte 024 pentru atenția asupra comportamentului suicidar" [translation of the Romanian version: "The Ministry of Health promotes the Hot line 024 for attention towards suicidal behaviour". The underlined words in the original translation and the translation of the Romanian version show that the meaning was completely changed in both cases. Second, both engines failed to be coherent when choosing the target terms throughout the same text, changing for no apparent reason since the same context was applicable. This is the case of "autocita-autocite-selfquote-selfappointment" (in the English corpus) and of several inconsistencies between the formal and informal way of addressing the reader (in the Romanian corpus). Third, we found several translation errors for examples of everyday language (e.g., related to smoking). Finally, the translation of metaphors (e.g., campaña 'Malos Humos' ['Smoking Is Bad' Campaign]; 'El Humo No Te Deja Ver' ['Smoke Blinds You']) has also been a challenge for both languages, especially for DeepL in the case of the English translations and for GT in the case of the Romanian translations.

Finally, we included the types of MT errors we found in the analysis of the DeepL and GT outputs for both language combinations and their frequency in Table 5:

Campaign (ES)	**Nº of words**	**Type of error**	**DeepL (EN)**	**GT (EN)**	**DeepL (RO)**	**GT (RO)**
1. Malos Humos	206	False meanings	2	2	4	6

		Syntactic and lexical calques and lexicalised metaphors	3	2	6	5
		Terminological improprieties			3	1
		Grammar mistakes				1
		Unnecessary foreign words or no translations				1
		Repetitions				
2. Alcohol	148	False meanings				3
		Syntactic and lexical calques and lexicalised metaphors	1	0	1	2
		Terminological improprieties			2	2
		Grammar mistakes				1
3. Conducta suicida	185	False meanings			1	1
		Syntactic and lexical calques and lexicalised metaphors		1		1
		Terminological improprieties	1	0	3	4
		Grammar mistakes				1
		Unnecessary foreign words or no translations				
		Repetitions				1

Table 5: Number of errors by type of text

Furthermore, Table 6 shows a comparative summary of the number and types of translation errors in both language combinations and using both translation engines. We found a variety of translation errors (nine types grouped in eight categories) of the list we established in section 2 and different results considering the two translation engines and the two language combinations involved. In this case, the Romanian translations had translation errors from all the categories included in Table 1 for both translation engines while the English translations had errors for six of the eight categories included. We also found more translation errors in the case of the Spanish-Romanian combination.

Type of error	**DeepL (EN)**	**GT (EN)**	**DeepL (RO)**	**GT (RO)**
False meanings	5	3	15	28
Syntactic and lexical calques and lexicalised metaphors	5	3	17	30
Terminological improprieties/imprecise expressions	3	0	20	26
Grammar mistakes	2	3	5	11
Omissions	2	1	1	1
Loans or no translation	0	0	2	7
Repetitions	1	0	1	2
Spelling mistakes	0	0	1	4

Table 6: Summary of the types of errors

Finally, we determined the percentage of translation errors considering the total number of words in the texts we analysed (539 words) and shown in Table 7. In general, they were below 3.5% in the English translations and not higher than 20% in the Romanian translations: we observed that the percentage was much higher in the Spanish-Romanian combination than in the case of the Spanish-English combination, as

	DeepL (EN)	GT (EN)	DeepL (RO)	GT (RO)
Total number of errors	18	10	62	109
%	3.33%	1.85%	11.5%	20%

Table 7. Percentage of translation errors

5 Discussions

Results show both similarities and differences considering the types of translation errors and the number of errors in the two language combinations involved considering the translations produced specifically in informative texts from health campaigns.

First, we observed that MT involved a variety of translation errors in these types of texts (see Tables 3, 4, and 5). The most common types of translation errors in both language combinations were false meanings, syntactic and lexical calques of the source language, terminological improprieties/lack of precision, grammar mistakes/errors, and the translation of lexicalised metaphors. We only found differences in the case of omissions and repetitions on the one hand (for the Spanish-English combination), and spelling mistakes and loans on the other hand (for the Spanish-Romanian combination).

Second, if we consider the results obtained for each engine and each language combination, we can underline several differences. The Romanian translations had more errors than the English translations in general. Moreover, in the case of the English translations, DeepL had more errors than GT, especially for terms with no context. On the opposite side, GT had more errors than DeepL in the Romanian translations, in some cases (false meanings and calques) with approx. 50% more errors in the GT output than in the DeepL output. In fact, all the Romanian texts included several examples of errors that hindered the correct transmission of the intended meaning: false meanings, syntactic and lexical calques of the source language, and terminological improprieties/imprecise expressions.

Therefore, although similar types of post-editing are required in both languages for adequacy purposes, there are important differences regarding the number of post-edits needed considering the two language combinations we compared. In this case, our results for informative texts from health campaigns suggest that the Romanian translations require much more post-editing than the English translations to achieve adequate results. Lastly, lack of coherence was also significant in both languages.

Third, despite the seriousness of the translation errors found in both language combinations and for both engines, the actual percentage of errors found was low considering the total number of words in the texts. This suggests that, although post-editing is required for adequate results that can be used to inform foreign users in a public setting, at least fragments of the texts translated are mostly functional. This means that the human effort required to apply post-editing strategies is generally low in both cases, especially for the English translations.

6 Conclusions

This study focused on the level of adequacy of the MT produced specifically in health campaigns' information and on the differences considering two language combinations.

The results we obtained helped us not only determine the level of adequacy of the translations for this type of texts but also reflect on their implications. Thus, if we refer to the transmission of information, the number of errors, and the seriousness of the translation errors, we could state that MT is only relatively adequate in the case of the English translations and not adequate in the case of Romanian translations. This verifies hypothesis 1, which suggests that MT alone is not adequate for the translation of informative health texts. Moreover, they showed that the number of translation errors and post-edits required is higher (and more varied) for the Spanish-Romanian combination, thus verifying hypothesis 2.

On the other hand, the results also made us reflect on the adequacy of using MT for informative texts combined with human post-editing as a strategy to benefit from the potential of MT while at the same time ensuring the quality of the public service translations. This strategy seems

to depend on the language combination and on the amount of time allotted for the task. Therefore, English translations have a higher level of adequacy and potential since we found fewer translation errors and less error variety, which involves less post-editing time. In turn, Romanian translations have a lower level of adequacy since we found more translation errors and more error variety, which involves more post-editing time.

In general, considering the low percentage of translation errors in both language combinations, one could argue that these outputs can be defined as functional, considering that the main messages intended in the original texts can be understood, which is the main purpose when using MT. It seems that the translation engines analysed could be used as a starting point in the translation process of informative healthcare texts with adequate post-editing strategies and the post-editing would not require a great amount of time from the translator. This process is easier when healthcare informative texts are involved since they tend to be repetitive and NMT can rely on a great amount of information available online especially in language combinations that involve English.

However, we still have to keep in mind that, for the time being, MT engines cannot fully render the natural-sounding language that the human translation produces. In fact, as our analysis shows, despite their continuous development, both engines still had difficulties finding adequate translations for idiomatic expressions, metaphors, and even in the translation of everyday language and consistency in both language combinations. Therefore, human post-editing is still essential to achieve a completely functional and understandable text that is adequate from the linguistic, social, and medical points of view. Ultimately, MT translation will continue to be researched, specifically within the PSIT context and will continue to improve its quality. This means that the results with PSIT texts such as the ones analysed in this research will keep improving.

References

Alarcón Navío, Esperanza. (2003). *Traducción automática versus traducción humana: tipología de errores*. In Muñoz Martín, Ricardo (Ed.), I AIETI. Actas del I Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación, 2 (pp. 721-738). AIETI.

Álvaro Aranda, Cristina. (2020). La traducción de materiales sanitarios sobre la COVID-19 para población inmigrante: análisis exploratorio de propuestas en España. *TRANS: Revista de Traductología, 24*, 455-468. https://doi.org/10.24310/TRANS.2020.v0i24.9762

Farrell, Michael. (2018). Raw output evaluator, a freeware tool for manually assessing raw outputs from different machine translation engines. In *Proceedings of the 40th Conference Translating and the Computer* (pp. 38-49). Asling: International Society for Advancement in Language Technology.

Gregorio Cano, Ana. (2017). Problemas de traducción, detección y descripción: un estudio longitudinal en la formación de traductores. *Revista digital de investigación en docencia universitaria, 11*(2), 26-49. https://doi.org/10.19083/ridu.11.552

Hurtado Albir, Amparo. (2011). *Traducción y traductología. Introducción a la traductología.* (5ºed.). Ediciones Cátedra.

INE (2022). Cifras de Población a 1 de julio de 2022. Estadística de Migraciones (EM). Primer semestre de 2022. Notas de Prensa. https://www.ine.es/prensa/cp_j2022_p.pdf

INE (1998-2022). Población extranjera por nacionalidad, comunidades, sexo y año. https://www.ine.es/jaxi/Tabla.htm?path=/t20/e245/p08/l0/&file=02005.px&L=0

Kerremans, Koen, Lázaro Gutiérrez, Raquel, Stengers, Helene, Cox, Antoon, & Rillof, Pascal. (2019). Technology use by public service interpreters and translators: The link between frequency of use and forms of prior training. *FITISPos International Journal*, *6*(1), 107-122.https://doi.org/10.37536/FITISPos-IJ.2019.6.1.211

Navaza, Bárbara, Estévez, Laura, & Serrano, Jordi, (2009). Saque la lengua, por favor. Panorama actual de la interpretación sanitaria en España. *Panace@, 10*(30), 141-156. https://dialnet.unirioja.es/servlet/articulo?codigo=3118253

Nord, Christiane. (2007). *Translating as a purposeful activity. Functionalist approaches explained. St. Jerome Publishing.*

Reiss, Katharina. (1983). Adequacy and equivalence in translation. *The Bible Translator*, *34*(3), 301-308. https://doi.org/10.1177/026009358303400301

Sánchez Ramos, María del Mar., & Rico Pérez, Celia. (2020). *CRISIS-MT: traducción y tecnología en la comunicación multilingüe para las ONG en situaciones de emergencia*. Vicerrectorado de Investigación y Transferencia de la Universidad de Alcalá.

Sixsmith, Jane, Doyle, Priscilla, D'Earth Maureen, & Barry, Margaret M. (2014). *Health communication and its role in the prevention and control of communicable diseases in Europe. Current evidence, practice and future developments.* European Centre for Disease Prevention and Control.

Valero-Garces, Carmen. (2018). Introduction. PSIT and Technology. Challenges in the Digital Age. *FITISPos International Journal*, *5*(1), 1-6. https://doi.org/10.37536/FITISPos-IJ.2018.5.1.185

Vázquez y del Arbol, Esther. (2008). Aspectos de la traducción técnica: Errores e incorreciones graves en la traducción de manuales de instrucciones alemanes e italianos al español. *Lebende Sprachen 53*(1), 2-6. https://doi.org/10.1515/LES.2008.2